EDA Process for Clustering Analysis

1. **Data Understanding**:
   Initially, we examined the structure of the customer and transaction datasets, understanding key features like CustomerID, Region, SignupDate, ProductID, ReviewCount, and Rating to assess their relevance for clustering.

2. **Data Cleaning**:
   We handled missing values and ensured that data types were appropriate (e.g., converting SignupDate to datetime format and ensuring CustomerID was treated as a string). Outliers or irrelevant columns were also addressed to prevent skewed results.

3. **Feature Engineering**:
   We created new features like DaysSinceSignup to measure customer activity and engagement, and scaled numerical features (e.g., ReviewCount, Rating) to ensure they were comparable across clusters.

4. **Exploratory Data Analysis**:
   Visualizations like histograms, box plots, and scatter plots helped identify patterns and distributions in customer signups, product ratings, and review counts. Correlations between features were also explored to guide the clustering process.

5. **Clustering and Evaluation**:
   We applied KMeans clustering on the cleaned and engineered features, followed by assessing cluster quality using the Davies-Bouldin Index and visualizing clusters through PCA to ensure meaningful groupings.